Journal of Biomolecular NMR, **14:** 149–155, 1999.
KLUWER/ESCOM

Refinement of the structure of protein–RNA complexes by residual dipolar coupling analysis

Peter Bayer*, Luca Varani & Gabriele Varani**
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Received 7 January 1999; Accepted 17 March 1999

Key words: protein-RNA, residual dipolar couplings, RNA structure, structure refinement

Abstract

The main limitation in NMR-determined structures of nucleic acids and their complexes with proteins derives from the elongated, non-globular nature of physiologically important DNA and RNA molecules. Since it is generally not possible to obtain long-range distance constraints between distinct regions of the structure, long-range properties such as bending or kinking at sites of protein recognition cannot be determined accurately nor precisely. Here we show that use of residual dipolar couplings in the refinement of the structure of a protein–RNA complex improves the definition of the long-range properties of the RNA. These features are often an important aspect of molecular recognition and biological function; therefore, their improved definition is of significant value in RNA structural biology.

Introduction

The main source of information in the determination of the three-dimensional structure of biomolecules by NMR is provided by distance and dihedral angle constraints. These constraints are local in nature, but suffice nonetheless to determine precise structures of globular proteins and their complexes. The local structure of RNA and DNA can also be determined to precision and accuracy comparable to those of proteins (Allain and Varani, 1997). However, nucleic acid structures are often more elongated than proteins. As a consequence, their global structures are generally poorly defined because there are no measurable long-range distance constraints connecting different regions of the same molecule (Varani et al., 1996; Allain and Varani, 1997). This is a major factor limiting the quality of NMR structures of DNA and RNA and of their complexes with proteins.

An important development in NMR-based structure determination has been the introduction of methods to measure the orientation of NH, CH or CC bond vectors relative to a reference axis system. This is best achieved by utilizing liquid crystalline solutions that can be ordered in the external magnetic field (Bax and Tjandra, 1997; Tjandra and Bax, 1997; Tjandra et al., 1997). Biomolecules can be oriented in these solutions and partial orientation gives raise to non-vanishing dipole-dipole interactions. The strength of the residual dipolar interaction depends on the degree of alignment (anisotropy) and on the direction of the bond vector with respect to the orientation axis, and therefore provides non-local structural information.

Residual dipolar couplings have been successfully applied in the refinement of protein structures (Tjandra and Bax, 1997; Tjandra et al., 1997), but they could have an even more significant impact in the study of RNA, DNA and of their complexes. In order to examine the impact of residual dipolar couplings on the quality of the structure of protein–RNA complexes, we re-examined the complex between human U1A protein and the polyadenylation regulatory RNA element within the mRNA coding for U1A itself. This complex represents a paradigm in protein–RNA recognition and in NMR studies of such systems (Allain et al., 1996, 1997). The structure was determined to

*Present address: Forschungsstelle für Enzymologie der Proteinfaltung, Weinberg 22, D-06210 Halle/Salle, Germany.
**To whom correspondence should be addressed. E-mail: gv1@mrc-lmb.cam.ac.uk

excellent precision from over 2500 distance and dihedral angle constraints (Allain et al., 1997; Howe et al., 1998). However, the orientation of the RNA double helical regions that emerge from the protein–RNA interface was not restricted experimentally and therefore was poorly defined in the complex. Here we show that measurement of residual dipolar couplings allows a more accurate definition of the long-range properties of the RNA in the complex, thereby alleviating the main limitation of NMR structures of protein–RNA complexes.

Materials and methods

Sample preparation

Human U1A protein (residues 2–102) was overexpressed in BLE21(DE3) *E. coli* utilizing a kanamycin resistance vector (Gerchman et al., 1994) and purified as described (Howe et al., 1998). Polyadenylation Inhibition Element (PIE) RNA was prepared by in vitro transcription using T7 RNA polymerase as described (Gubser and Varani, 1996). Isotopically labelled protein samples (100% ^{15}N and 100% ^{15}N-^{2}D) were prepared by growing *E. coli* with $^{15}NH_4Cl$ and D_2O (Aldrich). Isotopically labelled RNA (100% ^{13}C-^{15}N) was prepared by utilizing suitable labelled nucleotide triphosphates as precursor during enzymatic RNA synthesis; these were prepared by standard methods (Batey et al., 1995). After purification, RNA and protein samples were extensively dialyzed against the NMR buffer (10 mM Na-phosphate, pH 6–6.5 without added salt) and mixed in equimolar amounts. Typical concentrations were 0.15–0.3 mM in the liquid crystalline solution.

A liquid crystalline fluid composed of a mixture of 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) and 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) (Sanders and Schwonek, 1992) was used in the study of residual dipolar couplings in proteins (Bax and Tjandra, 1997; Tjandra and Bax, 1997; Ottiger and Bax, 1998). The same 1:3 (mol:mol) ratio of a 5.5% to 6.5% DHPC:DMPC mixture was used to measure residual dipolar couplings in the RNA–protein complex; the lipid solution was prepared essentially as reported (Ottiger and Bax, 1998). The solution containing the protein–RNA complex was added to the proper amount of a stock 15% lipid mixture. The quadrupolar splitting of the deuteron resonance line of the lock signal was measured to verify the amount of lipids in the liquid crystalline phase; this splitting was proportional to the percentage of lipids for all concentrations used in this study.

NMR spectroscopy

Residual dipolar couplings were extracted from ^{1}H-^{15}N or ^{1}H-^{13}C HSQC spectra recorded with WATERGATE solvent suppression with 2K (t_2) × 256 or 512 (t_1) real points on a Bruker DMX600 spectrometer. Processing was done with Felix230 (MSI) using sinebell square window functions shifted by $\pi/3$ or $\pi/2$ before Fourier transformation. HSQC spectra of the protein-RNA complexes were recorded at 38 °C both in aqueous buffered solution and in the liquid crystalline phase prepared as described above. At the relatively large molecular weight of the complexes under investigation (^{15}N T2 are 30–45 ms at 27 °C in water), aqueous buffers at 38 °C provided much more favorable lineshapes compared to the low temperature isotropic phase of the lipid solution. One-bond scalar couplings were measured either in t_1 or t_2 simply by not decoupling the appropriate heteronucleus during direct or indirect acquisition, as required. Couplings were measured in both dimensions for some samples in order to evaluate the uncertainty in the measurement. Residual dipolar couplings were then calculated according to:

$$D_{XH} = S_{XH311K,lipids} - S_{XH311K,buffer}$$

D_{XH} are the residual dipolar couplings and S_{XH} are the measured splittings between nuclei X and H (Figure 1). The precision of the measurement was ±1.2 Hz (backbone and imino NH); ±1.5 Hz (base CH) and ±2 Hz (sugar CH); these values refer to the experimental uncertainty in peak-to-peak separation in independent measurements of the same cross peak.

Preparation of residual dipolar coupling-based constraints

Use of residual dipolar couplings in structure refinement requires an evaluation of the anisotropy and the asymmetry of the system under investigation. This was done first by utilizing the 'powder distribution' of residual dipolar couplings (Figure 2A) (Clore et al., 1998). However, the diagonal components of the alignment tensor (d11, d22 and d33) can only be reliably estimated from the powder pattern if sufficient XH bond vectors can be monitored to sample every possible orientation in space. This condition is unlikely to be fulfilled for small proteins and, especially, nucleic acids. Therefore, we obtained the values of the

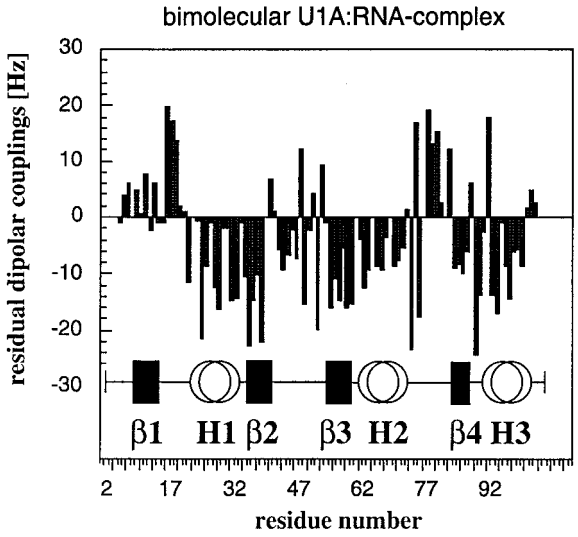

Figure 1. Residual dipolar couplings for backbone NHs in human U1A protein in complex with PIE RNA; the protein secondary structure is indicated at the bottom of the plot.

asymmetry by performing a set of structure calculations, monitoring the value of the target function (total energy, energy of residual dipolar couplings, etc.) that reports on the quality of the structure (Clore et al., 1998) (Figure 2B).

Structure calculation

Structure calculations were performed using X-PLOR 3.8 (Brünger, 1996) employing a specific subroutine to incorporate residual dipolar couplings (Tjandra et al., 1997b) in the simulated annealing and refinement protocol. Calculations were executed on a parallelized cluster of DEC Alfa workstations. Superpositions and structural statistics were done with Clusterpose (Diamond, 1995) and visualization was done with InsightII (MSI).

Structure calculations were based on two sets of constraints. A first set, identical to that used in the original structure determination (Allain et al., 1997; Howe et al., 1998), was used to test the performance of the refinement protocol. The second set contained 119 additional residual dipolar coupling constraints as shown in Table 1. Fifty structures derived from previously determined structures of the same complex were used to initiate the refinement protocol (Allain et al., 1997; Howe et al., 1998). Converged and non-converged structures were separated according to the agreement with experimental constraints and covalent geometry (Howe et al., 1998).

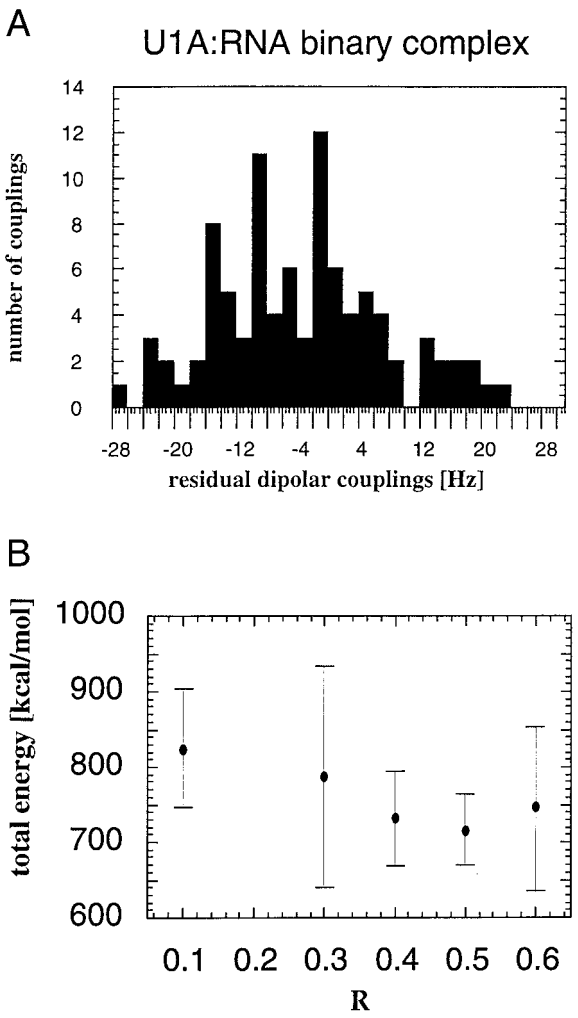

Figure 2. (A) 'Powder pattern' for residual dipolar couplings for backbone NHs in human U1A protein; (B) total energy of refined structures for different values of the anisotropy R.

Results and discussion

Residual dipolar couplings measured for amide NHs in human U1A(2–102) protein in complex with PIE RNA are shown in Figure 1. A total of 95 couplings were measured; 86 could be used as constraints for the structure calculation, whereas 9 were excluded due to ambiguity in resonance assignments in the liquid crystalline phase or mobility in the disordered N- and C-terminal ends of the protein. Using a $^{13}C/^{15}N$ labelled RNA sample, we could also extract 9 esidual dipolar coupling constants for imino NH pairs, 16 constraints for CH pairs within the bases and 9 constraints for CH pairs within the sugar. Mainly positive values were found for all imino NH bond vectors and for

Table 1. Experimental constraints and structural statistics

	Experimental constraints			
	NOE	H-bonds	Dihedrals	Residual J's
Protein	1710	42	0	86
RNA	591	25	110	33
Intermolecular	123	0	n.a.	n.a.
Structure statistics[a]		Set 1	Set 2	Allain et al.[b]
Protein backbone (8–96)		0.77±0.10	0.89±0.16	0.54±0.07
Protein heavy atoms (8–96)		1.24±0.13	1.30±0.14	0.93±0.05
RNA heavy atoms (20–49)		2.01±0.57	2.33±1.06	2.01±0.56
Protein binding site loop		0.90±0.15	1.16±0.35	0.86±0.12
Heavy atoms (8–96 U1A, 20–49)		1.90±0.37	2.18±0.69	2.05±0.54
All atoms		2.40±0.25	2.75±0.58	2.57±0.47

[a]Comparisons are based on the 12 structures of lowest energy obtained by each set of calculations; the mean deviation from ideal covalent geometry is as reported in other publications from our group (Allain et al., 1997; Allain and Varani, 1997; Howe et al., 1998).
[b]Refers to Allain et al. (1997) and Howe et al. (1998).

base aromatic CHs (data not shown). This immediately suggests that the three α-helices in the protein (which have negative residual dipolar couplings, Figure 1) are arranged approximately perpendicular to the RNA bases.

Nearly all observable imino protons could be used in this analysis, but only half of all base protons were sufficiently well resolved from other resonances to allow measurement of reliable dipolar coupling constants and very few sugar resonances could be resolved. In addition to spectral overlap, we encountered two major problems: increased sample aggregation in the liquid crystalline phase and the appearance of sharp cross peaks due to RNAs partially degraded by nucleases. It is very likely that the phospholipid preparation is contaminated with RNases that progressively degrade RNA (R. Gonzalez, personal communication). Although the present liquid crystalline media are adequate, alternative liquid crystalline phases (Hansen et al., 1998; Losonczi and Prestegard, 1998; Prosser et al., 1998; Wang et al., 1998) or alternative methods to obtain magnetic alignment (Beger et al., 1998) should be tested for obtaining more numerous and more precise residual dipolar couplings in RNA.

In order to utilize the data of Figure 1 for structure refinement, it was necessary to extract the molecular alignment tensor D. The anisotropy could be obtained from the 'powder pattern' of Figure 2A as described (Clore et al., 1998). Since this parameter tends to be underestimated by 15–20% by this procedure (Clore et al., 1998), the estimated value (−13.5 Hz) was increased to −15 Hz during the structure calculation. A critical uncertainty in deriving alignment tensor parameters from the powder pattern follows from the assumption that the distribution of NH bond vectors covers all possible directions in space. This is unlikely for a small protein such as U1A. The value of the asymmetry R was therefore obtained using a variational method (Clore et al., 1998). The average total energy of a family of 9 structures calculated for different values of R (0.1 to 0.6) is reported versus R in Figure 2B. These structures were calculated including all 2581 distance and dihedral angle constraints (Howe et al., 1998) and 86 residual dipolar couplings for amide NHs. The minimum in the plot provides a value for R=0.5, in good agreement with the value obtained from the powder pattern. It would be fruitless to use the powder pattern to find the value for D_a^{CH} in the RNA, since base NH and CH vectors in RNA and DNA are distributed far from randomly. This is very likely to remain a problem for RNA molecules of significant size (> 20 nucleotides), for which spectral overlap in the sugar region prevents quantitation of a sufficiently large number of residual dipolar couplings. For the anisotropy for CH bonds, we simply used a value twice that of the anisotropy for NH bonds, since

$$\gamma_{CH}/(r_{CH})^3 : \gamma_{NH}/(r_{NH})^3 \approx 2$$

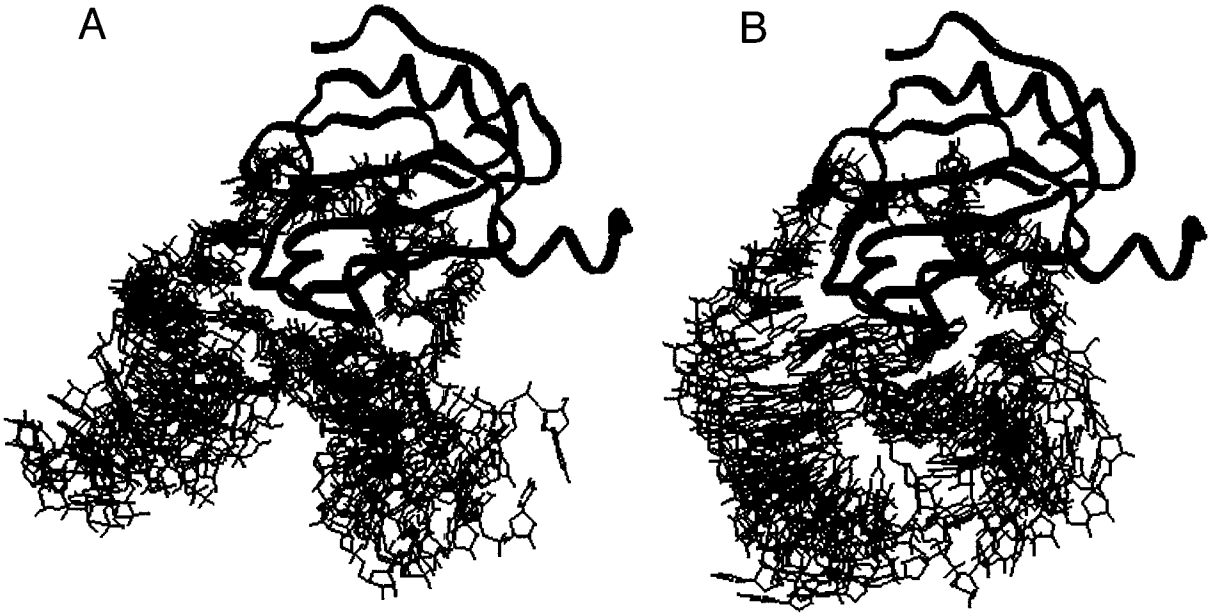

Figure 3. Superposition of low-energy structures refined with (A) or without (B) residual dipolar couplings.

Fifty starting coordinate sets derived from previously determined structures (Allain et al., 1997; Howe et al., 1998) were refined utilizing an X-PLOR-based protocol capable of incorporating constraints derived from residual dipolar couplings as a harmonic term into the target function (Tjandra et al., 1997):

$$E_{dipolar} = k_{dipolar}(\delta calc - \delta obs)^2$$

in addition to all other constraints present in a normal structure calculation. The force constants $k^{NH}_{dipolar}$ and $k^{CH}_{dipolar}$ were set to 0.55 kcal/mol and 0.35 kcal/mol, respectively (Tjandra et al., 1997; Clore et al., 1998). Two sets of calculations were performed using distinct sets of constraints, containing all experimental constraints previously reported, set 1, and all those constraints plus residual dipolar coupling constants, set 2. Energy profiles were used to separate converged and non-converged structures (Howe et al., 1998; Varani et al., 1996). A clusterfit algorithm was used to calculate the mean and standard deviation of the structures from the average structure (Diamond, 1995). Structural statistics are summarized in Table 1, where they are also compared with the results reported in the past utilizing a different computational protocol (Allain et al., 1997; Howe et al., 1998).

We first evaluated the performance of the refinement protocol, since it is well known that protocols optimized for protein structure refinement are not effective with RNA (Wimberly, 1992). Reassuringly, the precision of structures calculated using two different X-PLOR-based protocols was similar, overall rmsd's being 2.57 and 2.40 Å, respectively (Table 1). However, a more careful analysis indicates that sampling of conformational space by the protocol that allows inclusion of residual dipolar coupling constraints may not be as exhaustive as by the protocol optimized for RNA–protein complexes (Allain et al., 1996, 1997; Howe et al., 1998). As shown in Table 1, structural statistics are not as good when orientational constraints were included. This is particularly clear in two loops in the protein (residues 47–52 and 87–92) and the neighboring RNA nucleotides are less well defined in the new set of calculations. Protein side chains from these two regions become involved in numerous intermolecular interactions with RNA bases, resulting in extensive interpenetration of the two structures; the new protocol appears to reproduce these interactions less precisely. Table 1 indicates that inclusion of additional constraints does not necessarily lead to a more precisely determined structure, although this may reflect a limitation of the computational algorithm.

Structures calculated with orientational constraints derived from residual dipolar couplings had a total X-PLOR energy typically 100 kcal/mol higher when residual dipolar couplings were not included. Inclusion of residual dipolar couplings did not affect the precision of the structure to any significant extent.

This was true both at the global level (the overall precision was 2.40 versus 2.75 Å rmsd in the two cases) but also at the level of individual regions of the structures. The most significant difference from the inclusion of dipolar coupling was the separation of low energy structures into two well defined families differing mostly in the relative orientation of the double helical stems. Structures belonging to the first family have better agreement with the experimental data, including lower total energy and fewer violations of NOE- and dipolar couplings-derived constraints.

The relative position of the two double helical RNA stems is not well defined in structures calculated without long-range constraints derived from residual dipolar couplings (Allain et al., 1996). Short-range distance and dihedral angle constraints do not bear any information on long-range structural properties, and the lack of direct experimental information is properly reflected in a poor definition of these features. Inclusion of orientational constraints derived from residual dipolar couplings changes the relative position of the two double helical regions of the RNA (Figure 3). In the family of structures of lowest energy, there is a relatively wide angle between the axes defined by the double helical regions of the RNA. Thus, residual dipolar constraints appear to be a powerful driving force to define these long-range features of RNA structure.

Conclusions

We have explored the use of residual dipolar coupling constants in the refinement of the structure of a protein–RNA complex. The complex had been determined utilizing over 2500 NOE, dihedral and hydrogen bonding constraints (Allain et al., 1996, 1997; Howe et al., 1998). Addition of 119 constraints derived from residual dipolar couplings (86 for the protein and 33 for the RNA) changes the global conformation of the RNA element significantly. The relative orientation of the two double helical stems becomes much more open upon inclusion of these additional constraints (Figure 3). Since the new structures are consistent with a larger set of experimental data, they are likely to be more accurate than the original set of structures. Although precision was not improved upon inclusion of the present set of residual dipolar coupling constants, it is possible that including a larger number of residual dipolar couplings and an improved computational algorithm would lead to a significant improvement in precision as well. Thus, use of residual dipolar coupling constants reduces one of the most significant limitations of NMR-based structure determination of nucleic acids, the impossibility to define long-range properties of these elongated structures (Allain and Varani, 1997). This improvement can be of significant biological value, since proteins that bind RNA often change its conformation. These changes can lead to long-range effects that could affect biological events at distance; improved definition of these features provides insight into these important biological processes.

Acknowledgements

It is a pleasure to thank Dr. Fréderic Allain (UCLA) and Dr. Andres Ramos for advice and suggestions on RNA structure calculation; Dr. David Neuhaus for numerous discussions on all aspects of this project; Drs. Ad Bax and Nico Tjandra (NIH, Bethesda) provided invaluable advice in different stages of the project and were exceptionally collegial throughout. P.B. and L.V. were supported by post-doctoral and graduate fellowships, respectively, from the European Union.

References

Allain, F.H.-T., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D. and Varani, G. (1996) *Nature*, **380**, 646–650.

Allain, F.H.-T., Howe, P.W.A., Neuhaus, D. and Varani, G. (1997) *EMBO J.*, **16**, 5764–5774.

Allain, F. H.-T. and Varani, G. (1997) *J. Mol. Biol.*, **267**, 338–351.

Batey, R.T., Battiste, J. L. and Williamson, J.R. (1995) *Methods Enzymol.*, **261**, 300–322.

Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.

Beger, R.D., Marathias, V.M., Volkman, B.F. and Bolton, P.H. (1998) *J. Magn. Reson.*, **135**, 256–259.

Brünger, A.T. (1996) *X-PLOR version 3.854*, Yale University, New Haven, CT.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998) *J. Magn. Reson.*, **131**, 159–162.

Diamond, R. (1995) *Acta Crystallogr.*, **D 51**, 127–135.

Gerchman, S.E., Graziano, V. and Ramakrishnan, V. (1994) *Protein Expr. Purif.*, **5**, 242–251.

Gubser, C.C. and Varani, G. (1996) *Biochemistry*, **35**, 2253–2267.

Hansen, M.R., Rance, M. and Pardi, A. (1998) *J. Am. Chem. Soc.*, **120**, 11210–11211.

Howe, P.W.A., Allain, F.H.-T., Varani, G. and Neuhaus, D. (1998) *J. Biomol. NMR*, **11**, 59–84.

Losonczi, J.A. and Prestegard, J.H. (1998) *J. Biomol. NMR*, **12**, 447–451.
Ottiger, M. and Bax, A. (1998) *J. Biomol. NMR*, **12**, 361–372.
Prosser, R.S., Losonczi, J.A. and Shiyanovskaya, I.V. (1998) *J. Am. Chem. Soc.*, **120**, 11010–11011.
Sanders, C.R.I. and Schwonek, J.P. (1992) *Biochemistry*, **31**, 8898–8905.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Varani, G., Aboul-ela, F. and Allain, F.H.-T. (1996) *Prog. NMR Spectrosc.*, **29**, 51–127.
Wang, H., Eberstadt, M., Olejniczak, E.T., Meadows, R.P. and Fesik, S.W. (1998) *J. Biomol. NMR*, **12**, 443–446.
Wimberly, B.T. (1992) Ph.D. Thesis, University of California, Berkeley, CA.